

6. Management and Governance

AWS Auto Scaling

O Amazon EC2 Auto Scaling ajuda a manter a disponibilidade das aplicações e permite adicionar ou remover instâncias do EC2 automaticamente de acordo com as condições que você definir.

É possível usar os recursos de gerenciamento de frota do EC2 Auto Scaling para manter a integridade e a disponibilidade da sua frota.

Você também pode usar os recursos de escalabilidade dinâmica e preditiva do EC2 Auto Scaling para adicionar ou remover instâncias do EC2.

A **escalabilidade dinâmica** responde às mudanças na demanda e a **escalabilidade preditiva** agenda automaticamente o número certo de instâncias do EC2 com base na demanda prevista.

A escalabilidade dinâmica e a escalabilidade preditiva podem ser usadas juntas para escalas mais rápidas.

Gerenciamento de frota

Independentemente de executar uma instância do Amazon EC2, ou milhares, você pode usar o Amazon EC2 Auto Scaling para detectar instâncias deficientes e aplicativos não saudáveis do Amazon EC2 e substituir essas instâncias sem intervenção manual. Isso garante que seu aplicativo receba a capacidade computacional esperada. O Amazon EC2 Auto Scaling realizará três funções principais para automatizar o gerenciamento de frota para instâncias do EC2:

- Monitorar a integridade das instâncias em execução
O Amazon EC2 Auto Scaling garante que a sua aplicação esteja apta a receber tráfego e que as instâncias do EC2 estejam funcionando corretamente. O Amazon EC2 Auto Scaling realiza verificações periódicas de integridade para identificar quaisquer instâncias que não estejam íntegras.
- Substituir automaticamente as instâncias com falha
Quando uma instância com falha é reprovada em uma verificação de integridade, o Amazon EC2 Auto Scaling a encerra automaticamente e a substitui por uma nova. Isso significa que não é necessário responder manualmente quando uma instância precisa ser substituída.
- Balancear a capacidade entre as zonas de disponibilidades
O Amazon EC2 Auto Scaling pode balancear automaticamente instâncias entre zonas e sempre executa novas instâncias de modo que elas sejam balanceadas entre zonas o mais uniformemente possível em toda a frota.

Escalabilidade agendada (Scheduled scaling)

A escalabilidade com base em agendamento permite dimensionar seu aplicativo antes das alterações de carga conhecidas. Por exemplo, toda semana o tráfego para seu aplicativo da Web começa a aumentar na quarta-feira, permanece alto na quinta-feira e começa a diminuir na sexta-feira. Você pode planejar suas atividades de escalabilidade com base nos padrões de tráfego conhecidos de seu aplicativo da Web.

Escalabilidade dinâmica (Step Scaling vs Simple Scaling Policies vs Target Tracking Policies)

O Amazon EC2 Auto Scaling permite acompanhar de perto a curva de demanda por aplicativos, reduzindo a necessidade de provisionar antecipada e manualmente a capacidade do Amazon EC2. Por exemplo, é possível usar as políticas de escalabilidade de rastreamento de destino para selecionar uma métrica de carregamento para aplicações, como a utilização de CPU. Também existe a possibilidade de definir um valor de destino usando a nova métrica "Request Count Per Target" do Application Load Balancer, que é uma opção de balanceamento de carga do serviço de balanceamento de carga elástico. Depois disso, o Amazon EC2 Auto Scaling ajustará automaticamente o número de instâncias do EC2, conforme necessário para manter sua meta.

Step Scaling vs Simple Scaling Policies vs Target Tracking Policies

O EC2 Auto Scaling da Amazon fornece uma maneira eficaz de garantir que sua infraestrutura seja capaz de responder dinamicamente às mudanças nas demandas dos usuários. Por exemplo, para acomodar um aumento repentino de tráfego em seu aplicativo da web, você pode definir seu grupo Auto Scaling para adicionar automaticamente mais instâncias. E quando o tráfego estiver baixo, faça com que ele reduza automaticamente o número de instâncias. Esta é uma solução econômica, uma vez que provisiona instâncias do EC2 apenas quando você precisa delas. O EC2 Auto Scaling fornece várias políticas de escalonamento dinâmico para controlar os eventos de escalonamento interno e externo. Nos tópicos abaixo, discutiremos as diferenças entre uma política de dimensionamento simples, uma política de dimensionamento de etapas e uma política de rastreamento de destino. E mostraremos como criar um grupo de Auto Scaling com a política de step scaling aplicada.

Simple Scaling

O dimensionamento simples depende de uma métrica como base para o dimensionamento. Por exemplo, você pode definir um alarme do CloudWatch para ter um limite de **utilização da CPU** de 80% e, em seguida, definir a política de dimensionamento para adicionar 20% a mais de capacidade ao seu grupo Auto Scaling, iniciando novas instâncias. Da mesma forma, você também pode definir um alarme do CloudWatch para ter um limite de utilização da CPU de 30%. Quando o limite for atingido, o grupo Auto Scaling removerá 20% de sua capacidade encerrando as instâncias EC2. Quando o EC2 Auto Scaling foi introduzido pela primeira vez, esta era a única política de dimensionamento compatível. Ele não fornece nenhum controle refinado para aumentar e diminuir a escala.

Target Tracking

Com as políticas de escala de rastreamento de destino, você escolhe uma métrica de escala e define um valor de destino. O Application Auto Scaling cria e gerencia os alarmes do CloudWatch que acionam a política de dimensionamento e calcula o ajuste de dimensionamento com base na métrica e no valor alvo. A política de dimensionamento adiciona ou remove a capacidade conforme necessário para manter a métrica no valor de destino especificado ou próximo a ele. Além de manter a métrica próxima ao valor de destino, uma política de escala de rastreamento de destino também se ajusta às mudanças na métrica devido a uma mudança no padrão de carga.

A política de rastreamento de destino permite especificar uma métrica de escala e um valor de métrica que seu grupo de escala automática deve manter em todos os momentos. Digamos, por exemplo, que sua métrica de escalonamento é a utilização **média da CPU** de suas instâncias de escalonamento automático EC2 e que sua média deve ser sempre de 80%. Quando o CloudWatch detecta que a utilização média da CPU está além de 80%, ele aciona sua política de rastreamento de destino para dimensionar o grupo de escalonamento automático para atender a esta utilização de destino. Depois que tudo estiver resolvido e a utilização média da CPU ficar abaixo de 80%, outra escala em ação entrará em ação e reduzirá o número de instâncias de escalonamento automático em seu grupo de escalonamento automático. Com as políticas de rastreamento de destino, seu grupo de dimensionamento automático sempre estará executando em uma capacidade que é definida por sua métrica de dimensionamento e valor de métrica.

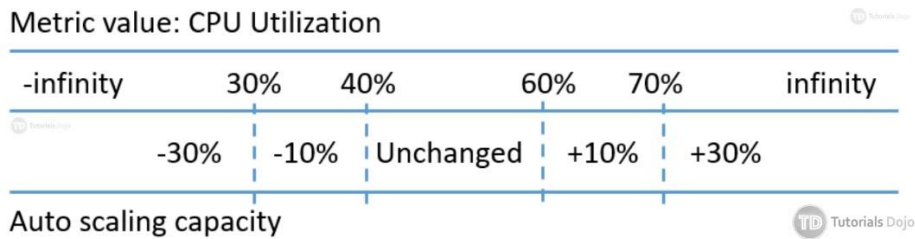
Uma limitação - este tipo de política pressupõe que deve escalar horizontalmente seu grupo de Auto Scaling quando a métrica especificada estiver acima do valor alvo. Você não pode usar uma política de escala de rastreamento de destino para dimensionar seu grupo de Auto Scaling quando a métrica especificada está abaixo do valor de destino. Além disso, o grupo Auto Scaling escala proporcionalmente à métrica o mais rápido possível, mas escala mais gradualmente. Por último, você pode usar métricas predefinidas da AWS para sua política de rastreamento de destino ou pode usar outras métricas CloudWatch disponíveis (nativas e personalizadas). As métricas predefinidas incluem o seguinte:

- ASGAverageCPUUtilization – Average CPU utilization of the Auto Scaling group.
- ASGAverageNetworkIn – Average number of bytes received on all network interfaces by the Auto Scaling group.
- ASGAverageNetworkOut – Average number of bytes sent out on all network interfaces by the Auto Scaling group.

- **ALBRequestCountPerTarget** – If the auto scaling group is associated with an ALB target group, this is the number of requests completed per target in the target group.

Step Scaling

O escalonamento de etapas melhora ainda mais os recursos de escalonamento simples. A escala de etapas aplica “ajustes de etapas”, o que significa que você pode definir várias ações para variar a escala, dependendo do tamanho da violação do alarme. Quando um evento de dimensionamento ocorre no dimensionamento simples, a política deve esperar que as verificações de integridade sejam concluídas e o resfriamento expire antes de responder a um alarme adicional. Isso causa um atraso no aumento da capacidade, especialmente quando há um aumento repentino de tráfego em seu aplicativo. Com o escalonamento em etapas, a política pode continuar a responder a alarmes adicionais mesmo no meio do evento de escalonamento. Aqui está um exemplo que mostra como funciona o escalonamento de etapas:



Neste exemplo, o grupo Auto Scaling mantém seu tamanho quando a utilização da CPU está entre 40% e 60%. Quando a utilização da CPU é maior ou igual a 60%, mas menor que 70%, o grupo Auto Scaling aumenta sua capacidade em 10% adicionais. Quando a utilização for superior a 70%, outra etapa de escalonamento é realizada e a capacidade é aumentada em 30% adicionais. Por outro lado, quando a utilização geral da CPU é menor ou igual a 40%, mas maior que 30%, o grupo Auto Scaling diminui a capacidade em 10%. E se a utilização cair ainda mais abaixo de 30%, o grupo Auto Scaling remove 30% da capacidade atual.

Isso fornece efetivamente várias etapas nas políticas de dimensionamento que podem ser usadas para ajustar a resposta do grupo do Auto Scaling à mudança dinâmica da carga de trabalho.

Se sua métrica de escala for uma métrica de utilização que aumenta ou diminui proporcionalmente à capacidade do destino escalonável, recomendamos que você use uma política de escala de rastreamento de destino.

Scenarios for termination policy use

The following sections describe the scenarios in which Amazon EC2 Auto Scaling uses termination policies.

- Scale-in events
- Instance refreshes
- Availability Zone rebalancing

Scale-in events

A scale-in event occurs when there is a new value for the desired capacity of an Auto Scaling group that is lower than the current capacity of the group.

Se você não atribuiu uma política de encerramento específica ao grupo, o Amazon EC2 Auto Scaling usa a política de encerramento padrão. Ele seleciona a Zona de disponibilidade com duas instâncias e encerra a instância que foi iniciada a partir do modelo de inicialização ou configuração de inicialização mais antigo. Se as instâncias foram

iniciadas a partir do mesmo modelo ou configuração de inicialização, o Amazon EC2 Auto Scaling seleciona a instância que está mais próxima da próxima hora de faturamento e a encerra.

Launch Configuration

Important: We strongly recommend that you do not use launch configurations. They do not provide full functionality for Amazon EC2 Auto Scaling or Amazon EC2. We provide information about launch configurations for customers who have not yet migrated from launch configurations to launch templates.

A *launch configuration* is an instance configuration template that an Auto Scaling group uses to launch EC2 instances. When you create a launch configuration, you specify information for the instances. Include the ID of the Amazon Machine Image (AMI), the instance type, a key pair, one or more security groups, and a block device mapping. If you've launched an EC2 instance before, you specified the same information in order to launch the instance.

You can specify your launch configuration with multiple Auto Scaling groups. However, you can only specify one launch configuration for an Auto Scaling group at a time, and you can't modify a launch configuration after you've created it. To change the launch configuration for an Auto Scaling group, you must create a launch configuration and then update your Auto Scaling group with it.

Keep in mind that whenever you create an Auto Scaling group, you must specify a launch configuration, a launch template, or an EC2 instance. When you create an Auto Scaling group using an EC2 instance, Amazon EC2 Auto Scaling automatically creates a launch configuration for you and associates it with the Auto Scaling group. Alternatively, if you are using launch templates, you can specify a launch template instead of a launch configuration or an EC2 instance.

Launch Template

A launch template is similar to a launch configuration, in that it specifies instance configuration information. Included are the ID of the Amazon Machine Image (AMI), the instance type, a key pair, security groups, and the other parameters that you use to launch EC2 instances. However, defining a launch template instead of a launch configuration allows you to have multiple versions of a template. With versioning, you can create a subset of the full set of parameters and then reuse it to create other templates or template versions. For example, you can create a default template that defines common configuration parameters and allow the other parameters to be specified as part of another version of the same template.

We recommend that you create Auto Scaling groups from launch templates to ensure that you're accessing the latest features and improvements. If you plan to continue to use launch configurations with Amazon EC2 Auto Scaling, be aware that not all Auto Scaling group features are available. For example, you cannot create an Auto Scaling group that launches both Spot and On-Demand Instances or that specifies multiple instance types or multiple launch templates. You must use a launch template to configure these features. For more information, see [Auto Scaling groups with multiple instance types and purchase options](#).

In addition to the features of Amazon EC2 Auto Scaling that you can configure by using launch templates, launch templates enable you to use newer features of Amazon EC2. This includes the current generation of EBS provisioned IOPS volumes (io2), EBS volume tagging, T2 Unlimited instances, elastic inference, and Dedicated Hosts, to name a few. Dedicated Hosts are physical servers with EC2 instance capacity that are dedicated to your use. While Amazon

EC2 Dedicated Instances also run on dedicated hardware, the advantage of using Dedicated Hosts over Dedicated Instances is that you can bring eligible software licenses from external vendors and use them on EC2 instances.

If you currently use launch configurations, you can specify a launch template when you update an Auto Scaling group that was created using a launch configuration.

To create a launch template to use with an Auto Scaling group, create the template from scratch, create a new version of an existing template, or copy the parameters from a launch configuration, running instance, or other template.

Auto Scaling Lifecycle Hooks

Lifecycle hooks allow you to control what happens when your Amazon EC2 instances are launched and terminated as you scale out and in. For example, you might download and install software when an instance is launching, and archive instance log files in Amazon Simple Storage Service (S3) when an instance is terminating.

Adding Lifecycle Hooks to the Auto Scaling group puts the instance into a wait state before termination. During this wait state, you can perform custom activities to retrieve critical operational data from a stateful instance.

Adicionar Lifecycle Hooks ao grupo Auto Scaling coloca a instância em um estado de espera antes do encerramento. Durante esse estado de espera, você pode executar atividades customizadas para recuperar dados operacionais críticos de uma instância com estado. O período de espera padrão é de 1 hora.

Cooldown period

o período de resfriamento ajuda a garantir que o grupo do Auto Scaling não inicie ou encerre instâncias adicionais antes que a atividade de dimensionamento anterior entre em vigor. Você pode configurar o período de tempo com base no período de aquecimento da instância ou em outras necessidades do aplicativo.

Health check grace period

frequentemente, uma instância do Auto Scaling que acabou de entrar em serviço precisa ser aquecida antes de passar na verificação de integridade. O Amazon EC2 Auto Scaling espera até que o período de carência da verificação de integridade termine antes de verificar o status de integridade da instância.

Metric type for Auto Scaling Group policy

Create Scaling policy

Name:

Metric type:

Application Load Balancer Request Count Per Target

Average CPU Utilization

Average Network In (Bytes)

Average Network Out (Bytes)

Target value:

Instances need:

300

seconds to warm up after scaling

Disable scale-in:

☐

[Create a simple scaling policy](#)

[Create a scaling policy with steps](#)

Create Alarm

You can use CloudWatch alarms to be notified automatically whenever metric data reaches a level that you specify.

To edit an alarm, first choose whom to notify and then define when the notification should be sent.

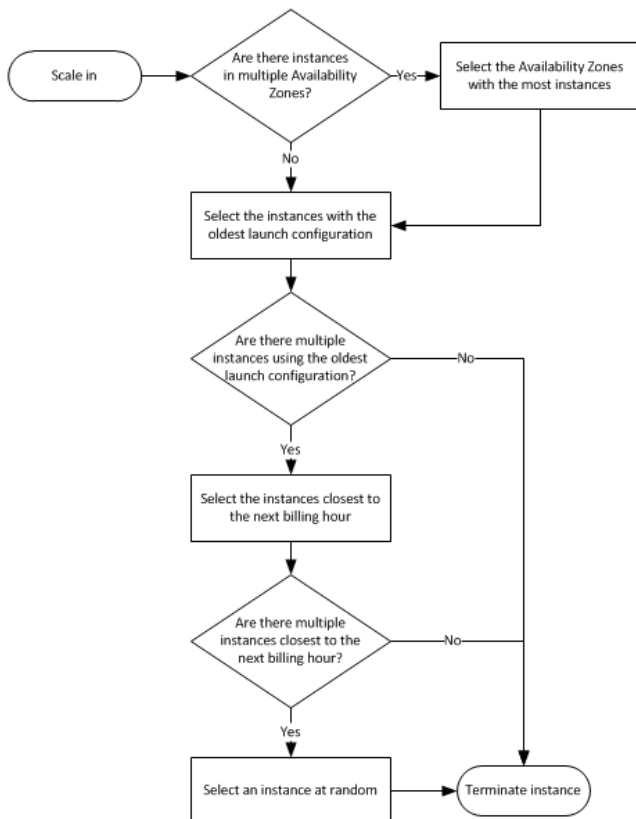
☒ Send a notification to: dynamodb (arn:aws:lambda:us-east-1:91) [create topic](#)

Whenever: Average of **CPU Utilization**
Is: >=
For at least: 1 consecutive

- CPU Utilization
- Disk Reads
- Disk Read Operations
- Disk Writes
- Disk Write Operations
- Network In
- Network Out

Scale in

The following flow diagram illustrates how the default termination policy works.



Pontos a se considerar

In most cases, step scaling policies are a better choice than simple scaling policies, even if you have only a single scaling adjustment.

You can have multiple target tracking scaling policies for an Auto Scaling group, provided that each of them uses a different metric. You can also have multiple scaling policies in force at the same time.

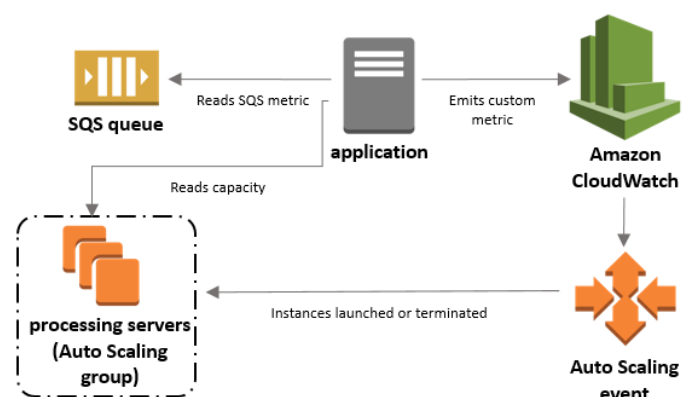
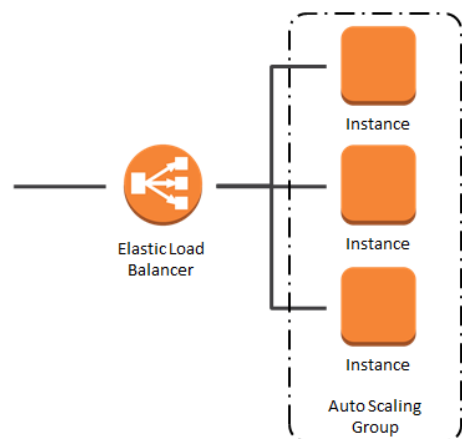
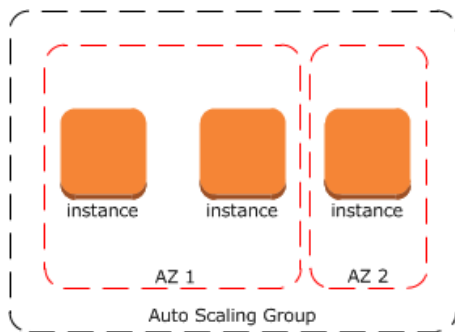
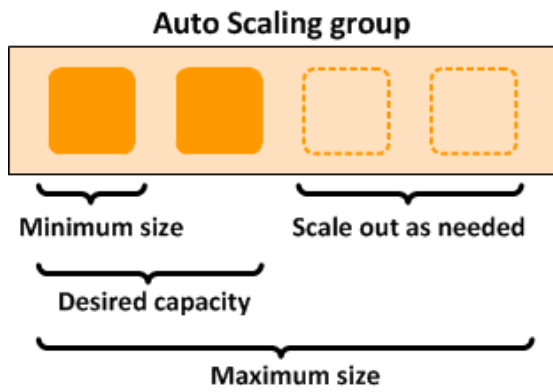
CloudWatch alarms associated with your target tracking scaling policies are deleted automatically when you delete the scaling policies.

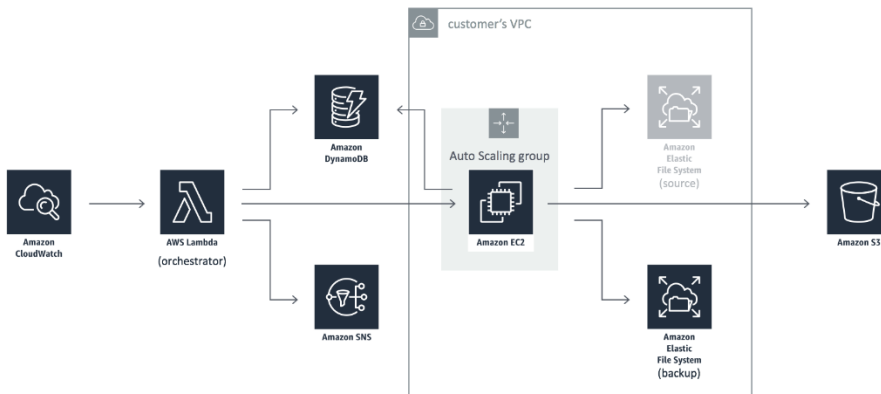
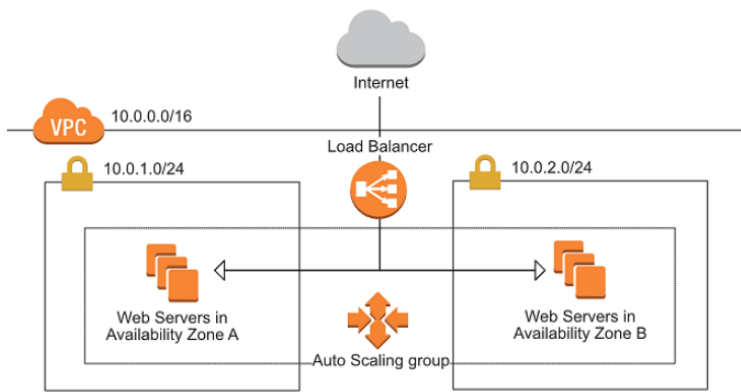
The gaps between the target value and the actual metric data points prevents you from adding an insufficient number of instances or removing too many instances.

Links Úteis

<https://docs.aws.amazon.com/autoscaling/ec2/userguide/lifecycle-hooks.html>

Algumas Imagens





➤ Pontos de Atenção

1. Um aplicativo é hospedado em um grupo de Auto Scaling de instâncias EC2. Para melhorar o processo de monitoramento, você deve configurar a capacidade atual para aumentar ou diminuir com base em um conjunto de ajustes de escala. Isso deve ser feito especificando as métricas de escalonamento e os valores de limite para os alarmes do CloudWatch que acionam o processo de escalonamento. Qual das opções a seguir é o tipo mais adequado de política de dimensionamento que você deve usar? R: Step scaling
2. Um conjunto de aplicativos da web é composto por vários grupos diferentes de Auto Scaling de instâncias EC2 que são configurados com configurações padrão e, em seguida, implantados em três Zonas de disponibilidade. Existe um Balanceador de Carga de Aplicativo que encaminha a solicitação ao respectivo grupo de destino no caminho da URL. A política de escalonamento foi acionada devido ao baixo número de tráfego de entrada para o aplicativo. Qual instância EC2 será a primeira a ser encerrada pelo grupo do Auto Scaling? R: A instância EC2 que pertence a um grupo Auto Scaling com a configuração de lançamento mais antiga
3. A tech company is currently using Auto Scaling for their web application. A new AMI now needs to be used for launching a fleet of EC2 instances. Which of the following changes needs to be done? R: Create a new launch configuration.
4. A commercial bank has designed their next generation online banking platform to use a distributed system architecture. As their Software Architect, you have to ensure that their architecture is highly scalable, yet still cost-effective. Which of the following will provide the most suitable solution for this scenario? R: Launch an Auto-Scaling group of EC2 instances to host your application services and an SQS queue. Include an Auto Scaling trigger to watch the SQS queue size which will either scale in or scale out the number of EC2 instances based on the queue.
5. You have an Auto Scaling group which is configured to launch new t2.micro EC2 instances when there is a significant load increase in the application. To cope with the demand, you now need to replace those instances

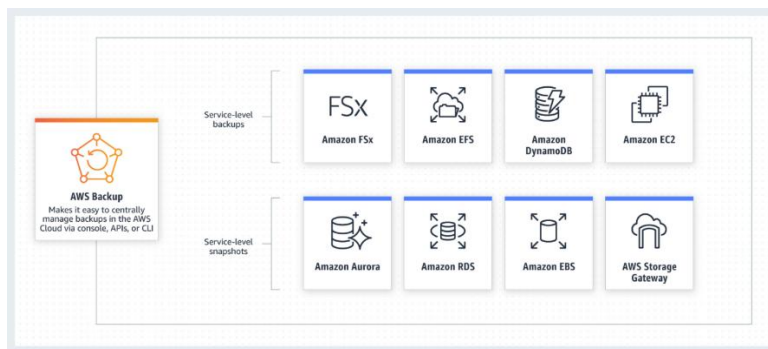
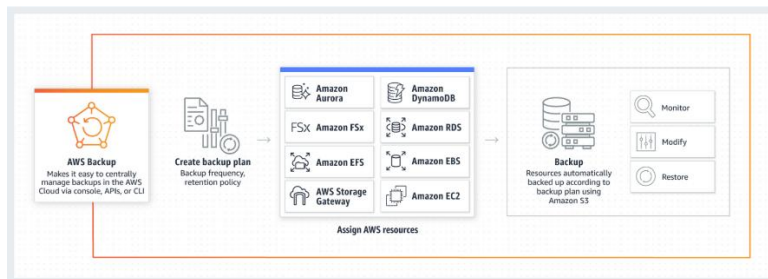
with a larger t2.2xlarge instance type. How would you implement this change? **R: Create a new launch configuration with the new instance type and update the Auto Scaling Group.**

6. An auto-scaling group of Linux EC2 instances is created with basic monitoring enabled in CloudWatch. You noticed that your application is slow so you asked one of your engineers to check all of your EC2 instances. After checking your instances, you noticed that the auto scaling group is not launching more instances as it should be, even though the servers already have high memory usage. What is the best solution that will fix this issue? R: Install CloudWatch monitoring scripts in the instances. Send custom metrics to CloudWatch which will trigger your Auto Scaling group to scale up.
7. Your organization had set up Auto Scaling for an EC2 instance. They intend to launch one additional new instance with same configuration automatically when the workload increases and shut it down automatically when the workload is back to normal. However, they have applied operating system patches to the main instance for security reasons and would like this to be reflected when the Auto Scaling group launches a new EC2 instance. What actions would you take in this scenario? R: Create an image out of main EC2 instance, create a new Launch Configuration with new image AMI ID, update Auto Scaling group with new Launch Configuration ID
8. Seu aplicativo consiste em um conjunto de instâncias EC2 geradas como parte de um grupo de escalonamento automático. Essas instâncias precisam acessar objetos em um bucket S3. Qual das alternativas a seguir é a abordagem ideal para garantir que esse acesso seja implementado? **R: Ensure that the launch configurations in Auto scaling group have an IAM Role to access S3 Objects.**

AWS Backup

Defina planos de backup, programe backups, automatize o gerenciamento da retenção de backups, monitore a atividade de backup de maneira centralizada e restaure os backups.

O Backup é realizado para os serviços ilustrados nas imagens:



Processo de Backup: Temos como exemplo a Amazon EC2, o AWS Backup faz o backup da/das instâncias gerando um AMI e copia o backup para outra Região conforme o agendamento configurado e conforme a regra estipulada

Agendamentos: é a frequência que o backup será realizado, podendo ser por hora, diariamente, semanalmente, mensalmente ou utilizar uma expressão Cron personalizada

Período de Retenção: O período de retenção do backup pode ser sempre, dias, semanas, meses ou anos

Tags e Recursos: é possível falar para o AWS Backup realizar o backup de todos os serviços que tem uma tag chamada por exemplo de "Backup=true"

Backup Consistente de Aplicações: Com o AWS Backup, você pode fazer backup e restaurar aplicativos Windows habilitados para VSS (Volume Shadow Copy Service) em execução em instâncias do Amazon EC2.

Observações: 1. É possível atribuir recursos a esse plano de backup após a criação dele. 2. É possível adicionar mais regras a esse plano de backup após a criação dele.

Definição de Preço

A cobrança pode ser realizada por GB/mês de armazenamento, por GB de Restauração e por transferência de dados entre regiões para cada serviço coberto pela AWS Backup. Exemplo:

Tipo de recurso	Armazenamento warm	Armazenamento inativo †
Backup do sistema de arquivos do Amazon EFS	0,05 USD por GB/mês	0,01 USD por GB/mês
Snapshot de volume do Amazon EBS	0,05 USD por GB/mês	n/d**
Snapshot do banco de dados do Amazon RDS	0,095 USD por GB/mês	n/d**
Snapshot do cluster do Amazon Aurora	0,021 USD por GB/mês	n/d**
Backup da tabela do Amazon DynamoDB	0,10 USD por GB/mês	n/d**
Backup do volume do AWS Storage Gateway	0,05 USD por GB/mês	n/d**
Backup do Amazon FSx for Windows File Server	0,05 USD por GB/mês	n/d**
Backup do Amazon FSx for Lustre	0,05 USD por GB/mês	n/d**

AWS CloudFormation

Infraestrutura Como Código

O AWS CloudFormation oferece uma linguagem comum para modelar e provisionar recursos de infraestrutura da AWS e terceirizados em um ambiente de nuvem

Ele permite usar linguagens de programação ou um arquivo de texto simples para modelar e provisionar de forma automática e segura todos os recursos necessários

Para um entendimento melhor: Tudo que é feito pelo SAM acaba se tornando um template do CloudFormation, ou seja, através de template é possível provisionar e configurar recursos da infraestrutura como criação de Lambdas, tabelas no DynamoDB, API Gateway, Filas, Tópicos e assim por diante

Ele provisiona recursos de forma segura e replicável, permitindo criar e recriar a sua infraestrutura e aplicativos, sem precisar executar ações manuais ou gravar scripts personalizados

Isso cria uma única fonte confiável para concentrar os seus recursos de infraestrutura, ou seja, é possível ter templates de ambiente de desenvolvimento e ambiente de produção, é possível também replicar os templates para outra Região ou outra Conta por exemplo

Porque isso é tão importante? Dessa forma, podemos replicar nosso ambiente com facilidade, sem nos preocuparmos se tudo foi recriado da forma como estava

Como Funciona?

Escrevemos nosso template em JSON ou YAML

Enviamos nosso template para o CloudFormation ou através do S3

Através do CloudFormation, CLI ou API, criamos uma Stack baseada no nosso template

O CloudFormation provisiona e configura os recursos especificados na Stack

Quando se deleta uma Stack do CloudFormation, automaticamente todos os serviços que foram criados pelo template também são deletados

Características

Infra como Código

- Controle
- Versões
- Visualizar as mudanças aplicadas

Custos

- Tags
- Custo estimado
- Scripts para criação/destruição de recursos

Definição de Preço

Paga-se pelos recursos da AWS (como instâncias do Amazon EC2, Load Balancers do Elastic Load Balancing etc.) criados com AWS CloudFormation como se tivesse criado os recursos manualmente

Links Úteis

https://docs.aws.amazon.com/pt_br/AWSCloudFormation/latest/UserGuide/aws-properties-sqs-queues.html

<https://docs.aws.amazon.com/autoscaling/ec2/userguide/asg-purchase-options.html>

<https://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/aws-properties-autoscaling-autoscalinggroup-instancesdistribution.html#cfn-autoscaling-autoscalinggroup-instancesdistribution-spotmaxprice>

➤ **Pontos de Atenção**

1. Você é um novo arquiteto de soluções em seu departamento e criou 7 modelos do CloudFormation. Cada modelo foi definido para um propósito específico. O que determina o custo de uso desses novos modelos CloudFormation? R: Os modelos do CloudFormation são gratuitos, mas você é cobrado pelos recursos subjacentes que ele cria.
2. The company you are working for has a set of AWS resources hosted in ap-northeast-1 region. You have been requested by your IT Manager to create a shell script which could create duplicate resources in another region in the event that ap-northeast-1 region fails. Which of the following AWS services could help fulfill this task? R: AWS CloudFormation
3. You use CloudFormation to create an Auto Scaling group for a web application. The application needs to be deployed in both non-production and production AWS accounts. You want to use Spot Instances in the non-production environment to save costs. Which of the following methods would you choose? R: In the CloudFormation template, use a parameter to set the **OnDemandPercentageAboveBaseCapacity** property. Set the parameter to be 0 in non-production and 100 in production.

AWS CloudTrail

O AWS CloudTrail é um serviço que possibilita governança, conformidade, auditoria operacional e auditoria de riscos em sua conta da AWS.

Com o CloudTrail, é possível registrar, monitorar continuamente e reter a atividade da conta relacionada às ações executadas na infraestrutura da AWS.

O CloudTrail disponibiliza o histórico de eventos da atividade da conta da AWS, inclusive ações executadas por meio do Console de Gerenciamento da AWS, dos AWS SDKs, das ferramentas da linha de comando e de outros Serviços da AWS.

Esse histórico de eventos simplifica a análise de segurança, o rastreamento de alterações de recursos e a solução de problemas.

Além disso, você pode usar o CloudTrail para detectar atividades incomuns em suas contas da AWS. Esses recursos ajudam a simplificar a análise operacional e a solução de problemas.

O AWS CloudTrail aumenta a visibilidade sobre as atividades de usuários e recursos ao registrar as ações e as chamadas de API do Console de Gerenciamento da AWS. Você pode identificar quais usuários e contas chamaram a AWS, a origem dos endereços IP de onde as chamadas foram feitas e quando as chamadas ocorreram.

O CloudTrail registra 2 tipos de eventos:

1. **Eventos de gerenciamento** que capturam ações do plano de controle, como criar ou excluir buckets do Amazon S3
2. **Eventos de dados** que capturam ações do plano de dados de alto volume, como leitura ou gravação de um objeto do Amazon S3.

O AWS CloudTrail registra eventos de gerenciamento nos produtos da AWS por padrão. Você pode exibir, pesquisar e fazer download (JSON ou CSV) do histórico de **90 dias** mais recente dos eventos de gerenciamento de sua conta gratuitamente usando o CloudTrail no console AWS ou a API de pesquisa da CLI da AWS.

Entregue eventos criando trilhas

Você pode entregar uma cópia de seus eventos de gerenciamento em andamento para o Amazon S3 gratuitamente criando trilhas. Isso permite que você armazene eventos dos últimos 90 dias no S3.

CloudWatch X CloudTrail

CloudWatch (* Performance):

- Computacional (CPU, Disk, Network)
- AutoScaling
- Elastic LoadBalancers
- Route 53
- CloudFront

CloudTrail (* Sistema de Câmeras / SysLog):

- Monitorar Usuários
- Comandos via CLI
- Aplicativos que acessam via API
- Inclusive eventos realizados no CloudWatch como por exemplo a criação de métricas

Links Úteis

<https://docs.aws.amazon.com/awscloudtrail/latest/userguide/how-cloudtrail-works.html>

➤ **Pontos de Atenção**

1. Você está trabalhando para uma empresa iniciante que possui recursos implantados na Nuvem AWS. Sua empresa agora está passando por um conjunto de auditorias programadas por uma empresa de auditoria externa para conformidade. Qual dos seguintes serviços disponíveis na AWS pode ser utilizado para ajudar a garantir que as informações corretas estejam presentes para fins de auditoria? R: AWS CloudTrail
2. You are managing an online platform which allows people to easily buy, sell, spend, and manage their cryptocurrency. To meet the strict IT audit requirements, each of the API calls on all of your AWS resources should be properly captured and recorded. You used CloudTrail in your VPC to help you in the compliance, operational auditing, and risk auditing of your AWS account. In this scenario, where does CloudTrail store all of the logs that it creates? R: Amazon S3
3. Uma empresa precisa monitorar a atividade da API para fins de auditoria em sua conta da AWS. Essa auditoria também seria realizada no futuro e deveria ser aplicada a todas as regiões. Como você projetaria sua solução para atender às necessidades presentes e futuras? R: Ensure one CloudTrail log is enabled for all regions. *Quando você cria uma trilha que se aplica a todas as regiões, o CloudTrail registra eventos em cada região e entrega os arquivos de log de eventos do CloudTrail para um bucket S3 que você especificar. Se uma região for adicionada após a criação de uma trilha que se aplica a todas as regiões, a nova região será incluída automaticamente e os eventos nessa região serão registrados.*

Amazon CloudWatch

O CloudWatch é um serviço de monitoramento para **Recursos** em nuvem AWS. Você pode usá-lo para coletar e rastrear métricas, coletar e monitorar arquivos de log, e definir alarmes

Você pode usá-lo para obter visibilidade sobre a utilização de recursos, a performance de aplicativos e o status operacional em todo o sistema, ou seja, é utilizado tanto para análise quanto para tomar ações efetivas

É possível usar essas percepções para reagir e manter seu aplicativo em execução tranquilamente. Ele pode monitorar recursos da AWS como instâncias do Amazon EC2, tabelas do DynamoDB, Funções Lambda, dentre outros

Você pode usar o CloudWatch para detectar comportamento anormal em seus ambientes, definir alarmes, visualizar logs e métricas lado a lado... executar ações automatizadas, resolver problemas e descobrir insights para manter seus aplicativos em perfeita execução. Exemplo 1: o Dynamo alcançou uma quantidade de leitura muito próxima do limite, nesse caso pode se disparar um alarme para avisar o administrador ou simplesmente tomar alguma ação. Exemplo 2: instâncias EC2 quando estiverem chegando próximo do limite o próprio CloudWatch iniciar o Auto Scaling

Simplicidade: É fácil monitorar seus recursos e aplicativos da AWS com o CloudWatch. Ele se integra nativamente com mais de 70 serviços da AWS, como AWS EC2, AWS DynamoDB, AWS S3, AWS ECS, AWS EKS e AWS Lambda, além de publicar automaticamente métricas detalhadas de um minuto e métricas personalizadas com até um segundo de granularidade

Podemos utilizar o CloudWatch para ambientes sem servidor (Serverless), como por exemplo para iniciar o AWS Lambda. Exemplo: uma função é startada todo dia ao meio dia, o recurso de CRON do CloudWatch pode ser usado nestes casos

Definição de Preço

Nível Gratuito: A maioria dos serviços da AWS (EC2, S3, Kinesis, etc.) fornece automaticamente métricas gratuitas para o CloudWatch. Muitos aplicativos conseguem operar dentro desses limites do nível gratuito

Métricas	Métricas de monitoramento básico (com frequência de 5 minutos)
	10 métricas de monitoramento detalhado (com frequência de 1 minuto)
	1 milhão de solicitações de API (não aplicável a GetMetricData e GetMetricWidgetImage)
Painel	3 painéis com até 50 métricas por mês
Alarmes	10 métricas de alarmes (não aplicáveis a alarmes de alta resolução)
Logs	5 GB de dados (consumo, armazenamento de arquivo e dados verificados pelas consultas do Logs Insights)
Eventos	Todos os eventos estão incluídos, exceto os personalizados
Contributor Insights	1 regra do Contributor Insights por mês
	Os primeiros um milhão de eventos de log que correspondem à regra por mês
Synthetics	100 execuções de canary por mês

Métricas

Existem 2 tipos de métricas, as Nativas e as Customizadas:

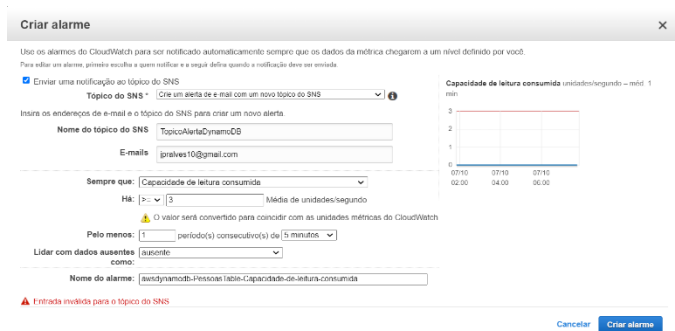
Nativas: As métricas variam conforme o tipo de recursos que está sendo monitorado, por exemplo, se for um tópico SNS, podemos ter métricas sobre quantidade mensagens por tópico, se for um Lambda, métricas sobre quantidade de memória, tempo de execução e assim por diante

Customizadas: Recursos como EC2 por exemplo, tem uma outra tratativa. O CloudWatch consegue saber quanto de CPU está sendo utilizando porém memoria ou disco não é possível, mesmo sabendo quanto de memória e disco está alocada para aquela instância. Para que o CloudWatch tenha acesso a essa info é possível configurar dentro da instância um agente ou scripts que monitorem e enviem a info para o CloudWatch

Dispare Eventos toda vez que um recurso exceder um limite ou houver uma anomalia, ou seja, é possível ativar um Lambda, disparar um e-mail dentre outras coisas

Alarmes

Use os alarmes do CloudWatch para ser notificado automaticamente sempre que os dados da métrica chegarem a um nível definido por você



Criar alarme

Use os alarmes do CloudWatch para ser notificado automaticamente sempre que os dados da métrica chegarem a um nível definido por você.

Para editar um alarme, primeiro escolha a ação notificar e a seguir defina quando a notificação deve ser enviada.

☒ Enviar uma notificação ao tópico do SNS

Tópico do SNS:

Insira os endereços de e-mail e o tópico do SNS para criar um novo alerta.

Nome do tópico do SNS:

E-mail:

Sempre que:

Hi: Média de unidades/segundo

Pelo menos: período(s) consecutivo(s) de 5 minutos

Lidar com dados ausentes como:

Nome do alarme:

O Amazon CloudWatch pode ajudar você a monitorar as cobranças na sua Fatura da AWS enviando alarmes por e-mail quando as cobranças excederem um limite definido por você.

Depois que você atualizar suas preferências no console de faturamento da conta, começará a receber métricas do Amazon CloudWatch que refletem suas cobranças da AWS acumuladas no mês. Em seguida, você pode criar um alerta de faturamento especificando um limite de gastos e um endereço de e-mail para notificar.

Tipo de Limite

Estático: Usar um valor como limite

Deteção de Anomalias: Usar um segmento como limite

Eventos/Regras

As regras roteiam eventos dos seus recursos da AWS para processamento por destinos selecionados. Você pode criar, editar e excluir regras.

CloudWatch Logs Agent

Important: This reference is for the older CloudWatch Logs agent, which is on the path to deprecation. We strongly recommend that you use the unified CloudWatch agent instead.

The CloudWatch Logs agent provides an automated way to send log data to CloudWatch Logs from Amazon EC2 instances. The agent includes the following components:

- A plug-in to the AWS CLI that pushes log data to CloudWatch Logs.
- A script (daemon) that initiates the process to push data to CloudWatch Logs.
- A cron job that ensures that the daemon is always running.

Links Úteis

<https://docs.aws.amazon.com/AmazonCloudWatch/latest/events/ScheduledEvents.html>

<https://docs.aws.amazon.com/AmazonCloudWatch/latest/logs/AgentReference.html>

<https://docs.aws.amazon.com/AmazonCloudWatch/latest/monitoring/UsingAlarmActions.html>

<https://aws.amazon.com/cloudwatch/faqs/>

➤ **Pontos de Atenção**

1. You are working as an AWS Engineer in a major telecommunications company in which you are tasked to make a network monitoring system. You launched an EC2 instance to host the monitoring system and used CloudWatch to monitor, store, and access the log files of your instance. Which of the following provides an automated way to send log data to CloudWatch Logs from your Amazon EC2 instance? **R: CloudWatch Logs agent**
2. You have a web application hosted in AWS cloud where the application logs are sent to Amazon CloudWatch. Lately, the web application has recently been encountering some errors which can be resolved simply by restarting the instance. What will you do to automatically restart the EC2 instances whenever the same application error occurs? R: First, look at the existing CloudWatch logs for keywords related to the application error to create a **custom metric**. Then, create a CloudWatch alarm for that custom metric which invokes an action to restart the EC2 instance.
3. You are a Solutions Architect for a large London-based software company. You are assigned to improve the performance and current processes of supporting the AWS resources in your VPC. Upon checking, you noticed that the Operations team does not have an automated way to monitor and resolve issues with their on-demand EC2 instances. What can be used to automatically monitor your EC2 instances and notify the Operations team for any incidents? R: AWS Cloudwatch
4. The company that you are working for has a highly available architecture consisting of an elastic load balancer and several EC2 instances configured with auto-scaling in three Availability Zones. You want to monitor your EC2 instances based on a particular metric, which is not readily available in CloudWatch. Which of the following is a custom metric in CloudWatch which you have to manually set up? **R: Memory Utilization of an EC2 instance**

AWS Config

O AWS Config é um serviço que permite acessar, auditar e avaliar as configurações dos recursos da AWS, ou seja, o que está em compliance e o que não está em compliance

O Config monitora e grava continuamente registros das configurações de recursos da AWS e lhe permite automatizar a avaliação das configurações registradas com base nas configurações desejadas

Com o Config, você pode analisar alterações feitas **nas configurações e relacionamentos entre os recursos da AWS**, aprofundar-se de forma detalhada no histórico de configuração de recursos e determinar a conformidade geral em relação às configurações especificadas em suas diretrizes internas

Dessa forma, você pode simplificar a auditoria de conformidade, a análise de segurança, o gerenciamento de mudanças e a solução de problemas operacionais

O AWS Config trabalha juntamente com o CloudTrail com a finalidade principal de mostrar tudo que aconteceu em um recurso, **e principalmente, faz checagens de segurança**

Por exemplo: AWS Config cheque todas as minhas instâncias EC2 e verifique se todas elas têm uma tag, se estão criptografadas, se estão com security group não liberando a porta 22 e assim por diante. Se alguma instância não estiver em compliance é possível criar uma ação nesse sentido, como por exemplo excluir a instância

Casos de Uso

Descoberta: descobrirá recursos que existem na sua conta, registrará suas configurações atuais e capturará qualquer alteração nessas configurações

Gerenciamento de Alterações: Quando seus recursos são criados, atualizados ou excluídos, o AWS Config transfere essas alterações na configuração para o Amazon Simple Notification Service (SNS), para que você seja notificado sobre todas as alterações de configuração. O AWS Config representa as relações entre recursos, para que você possa avaliar como uma alteração em um recurso pode afetar outros recursos

Auditoria e Conformidade Contínuas: O AWS Config foi projetado para ajudar você a avaliar a conformidade com as políticas e os padrões normativos internos, oferecendo visibilidade das configurações de seus recursos da AWS

Estrutura de Conformidade como Código: Você pode codificar seus requisitos de conformidade como regras do AWS Config e criar ações de correção usando documentos do AWS Systems Manager Automation e agrupá-los em um pacote de conformidade

Solução de Problemas: Usando o AWS Config, você pode solucionar problemas operacionais rapidamente identificando as alterações de configuração recentes em seus recursos

Análise de Segurança: As alterações feitas às configurações dos seus recursos podem acionar as notificações do Amazon Simple Notification Service (SNS), que são enviadas à equipe de segurança para análise e ações.

Definição de Preço

Itens de Configuração: Um item de configuração é um registro do estado de configuração de um recurso na sua conta da AWS

Avaliações de Regras: Uma avaliação de regra do AWS Config é uma avaliação do estado de conformidade de um recurso por uma regra do AWS Config em sua conta da AWS

Avaliações do Pacote: Avaliação do pacote de conformidade é a avaliação de um recurso por uma regra do AWS Config no pacote de conformidade

Com o AWS Config, você é cobrado com base no número de itens de configuração registrados, no número de avaliações de regras ativas do AWS Config e no número de avaliações do pacote de conformidade na conta

Resources

With AWS Config, you can do the following.

- Evaluate your AWS resource configurations for desired settings.
- Get a snapshot of the current configurations of the supported resources that are associated with your AWS account.
- Retrieve configurations of one or more resources that exist in your account.
- Retrieve historical configurations of one or more resources.
- Receive a notification whenever a resource is created, modified or deleted.
- View relationships between resources. For example, you might want to find all resources that use a particular security group.

➤ Pontos de Atenção

1. A multinational manufacturing company has multiple accounts in AWS to separate their various departments such as finance, human resources, engineering and many others. There is a requirement to ensure that certain access to services and actions are properly controlled to comply with the security policy of the company. As the Solutions Architect, which is the most suitable way to set up the multi-account AWS environment of the company? R: Use AWS Organizations and Service Control Policies to control services on each account.
2. You have a requirement to get a snapshot of the current configuration of resources in your AWS Account. Which service can be used for this purpose? R: AWS Config

Amazon EventBridge (Amazon CloudWatch Events)

Conceitos Importantes

Producer / Broker (EventBridge/Kafka) / Consumer

Producer: um producer, pode ser um app mobile ou a SDK da AWS por exemplo, que gera um evento ou vários eventos. Essas informações do evento normalmente são enviadas no formato JSON. Um evento de exemplo seria informar ao Broker a mudança de cidade do usuário daquele celular

Broker: o broker tem a capacidade de receber vários eventos e de forma assíncrona, gerencia-los através de microfilas. Ele pode ser consumido ou disparar uma trigger para o consumer

Consumer: o consumer (aplicação, lambda, SQS, SNS) baseado no tipo de informação recebida toma diversas ações, como por exemplo sugerir ao usuário por e-mail as melhores hamburguerias daquela cidade

CloudWatch Eventos x EventBridge

Basicamente as duas aplicações são a mesma coisa, pois no fim das contas tem a mesma arquitetura. No CloudWatch Eventos os serviços da AWS disparam eventos e baseado em regras pré determinadas dispara um evento para um consumer

Definição de Preço

Tamanho das cargas úteis: cada bloco de 64 KB de uma carga útil é cobrado como um evento (por exemplo, um evento com carga de 256 KB é cobrado como quatro solicitações).

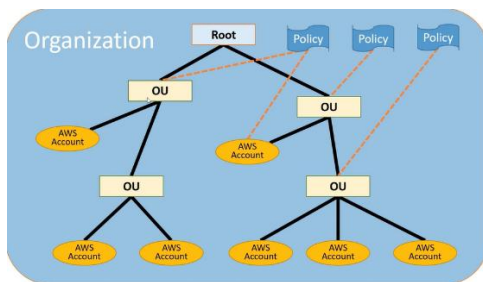
Eventos de serviços da AWS	Gratuito
Eventos personalizados	1,00 USD/milhões de eventos personalizados publicados
Eventos de terceiros (SaaS)	1,00 USD/milhões de eventos publicados
Eventos para outro barramento	1,00 USD/milhões de eventos publicados

AWS Organizations

O AWS Organizations é um serviço de gerenciamento de contas que permite consolidar várias contas da AWS em uma organização que você cria e gerência de maneira centralizada

Você pode usar o Organizations para aplicar políticas que oferecem a suas equipes a liberdade de criar com os recursos de que precisam, enquanto permanecem dentro dos limites de segurança que você definiu.

Ao organizar contas em Unidades Organizacionais (UOs), que são grupos de contas que atendem a uma aplicação ou serviço, você pode aplicar políticas de controle de serviço (SCPs) para criar limites de governança direcionados para suas UOs



Simplifique o gerenciamento de permissões com base no usuário para todos em sua organização com AWS Single Sign-On (SSO) e seu Active Directory. Você pode aplicar práticas de privilégios mínimos criando permissões personalizadas para categorias de trabalho. Você também pode controlar o acesso aos serviços da AWS aplicando políticas de controle de serviço (SCPs) a usuários, contas ou UOs

Proteção e Auditoria

Gerencie a auditoria em escala usando o **AWS CloudTrail** para criar um registro imutável de todos os eventos das contas. Você pode aplicar e monitorar os requisitos de backup com o **AWS Backup** ou definir centralmente seus critérios de configuração recomendados em recursos, regiões da AWS e contas com **AWS Config**



Vantagens da Conta Consolidada

- Uma fatura por conta da AWS
- Muito fácil de rastrear cobranças e alocar custos
- Desconto de preço por volume (por exemplo: o total de armazenamento de S3)
- Alocação de reservas de recursos entre as contas (foi reservado uma quantidade de 20 de um tipo de instância para um determinado time de desenvolvimento, porém esse time está usando somente 10, é possível redistribuir o restante para outra conta)

Melhores Práticas

- Sempre habilitar o MFA na conta Root e sempre usar senhas complexas na conta Root
- É recomendado que a conta Payer (pagante) seja usada apenas para “billing” e que nenhum outro serviço rode nela
- Usar sempre o SCP (Service Control Policies) para habilitar/desabilitar algum serviço de maneira global por OU's

Definição de Preço

O AWS Organizations é disponibilizado para todos os clientes da AWS sem cobranças adicionais

➤ Pontos de Atenção

1. You work in a start-up company as an AWS solutions architect. You create a new AWS Organization that includes a large amount of AWS accounts. You want to use a tool to log an event and trigger a notification whenever the administrator performs an action in the Organization. Which of the following AWS services would you use? R: AWS CloudWatch Events.
2. Uma empresa de mídia Firm_A usa infraestrutura AWS e tem uma presença global para sua programação de esportes e rede de transmissão. Ele usa a Organização AWS para gerenciar várias contas AWS. Recentemente, foi adquirido pela Firm_B que também usa a infraestrutura AWS. Firm_B também tem seus próprios conjuntos de contas AWS. Após a fusão, as contas da AWS de ambas as organizações precisam se fundir para criar e gerenciar políticas de forma mais eficaz. Como consultor da AWS, qual das etapas a seguir você sugeriria ao cliente para mover a conta mestre da empresa de mídia original para a organização usada pela entidade resultante da fusão? (Selecione TRÊS.) R: 1) Remove all member accounts from the organization in Firm_A. 2) Delete the organization in Firm_A. 3) Invite the Firm_A master account to join the new organization (Firm_B) as a member account.
3. Uma grande firma de arquitetos com sede em Cingapura está usando um balde Amazon S3 para salvar todos os desenhos de arquitetura. Esta empresa trabalha globalmente e várias contas AWS são criadas. Os usuários em todas essas contas acessam o bucket do Amazon S3 para desenhos arquitetônicos. Uma organização AWS é criada para várias contas da equipe de desenvolvimento. A equipe aplicou um SCP na conta mestre da Organização para permitir o acesso ao bucket S3. Qual das seguintes afirmações está correta? R: O SCP permitirá o acesso ao bucket do Amazon S3 para todas as contas dentro da organização, incluindo usuários root de cada conta e não para usuários fora desta organização. Service Control Policies will be applied to all users within member accounts including root accounts within each of the accounts. These policies are NOT applied to users outside of AWS Organizations.
4. Uma empresa iniciante está usando uma organização AWS para gerenciar políticas em suas contas de desenvolvimento e produção. A conta de desenvolvimento precisa de um host dedicado EC2. A conta de produção se inscreveu em um host dedicado EC2 para seu aplicativo, mas não o está usando no momento. O compartilhamento NÃO foi habilitado com a Organização AWS na AWS RAM. Qual das seguintes opções pode ser feita para compartilhar o host dedicado do Amazon EC2 da conta de produção para a conta de desenvolvimento? R: Create a resource share in the production account and accept the invitation in the development account. For accounts that are part of the AWS Organization, Resource sharing can be done on an individual account basis if resource sharing is not enabled at the AWS Organisation level. With this, resources are shared within accounts as external accounts & an invitation needs to be accepted between these accounts to start resource sharing.
5. Uma empresa global possui um banco de dados Amazon Aurora para armazenar uma grande quantidade de dados de clientes. O banco de dados é implantado em uma conta AWS de propriedade da equipe de desenvolvimento, e a conta AWS está dentro da Organização AWS A. Agora, o banco de dados precisa ser compartilhado com contas AWS em outra Organização AWS B. Qual das opções a seguir pode ser feita para alcançar o requerimento? R: Na conta Management AWS da Organização A, compartilhe o banco de dados com as contas AWS da AWS Organization B no **Resource Access Manager**.
6. Um cliente precisa de governança corporativa de TI e supervisão de custos de todos os recursos da AWS consumidos por suas divisões. Cada divisão tem sua própria conta da AWS e é necessário garantir que as políticas de segurança sejam mantidas no nível da conta. Como você pode conseguir isso? Escolha 2 respostas. R: **1) Use AWS organizations 2) Use Service control policies**

AWS Resource Access Manager

O AWS Resource Access Manager (RAM) ajuda você a compartilhar com segurança seus recursos entre contas AWS, dentro de sua organização ou unidades organizacionais (OUs) em **AWS Organizations** e com funções IAM e usuários IAM para tipos de recursos suportados.

Compartilhamento de Recursos com:

- Contas da AWS
- Unidades Organizacionais (OUs)
- Roles do IAM
- Usuários do IAM

Você pode usar AWS RAM para compartilhar transit gateways, sub-redes, configurações de licença do AWS License Manager, regras do Amazon Route 53 Resolver e mais tipos de recursos

Muitas organizações usam várias contas para criar isolamento administrativo ou de faturamento e para limitar o impacto de erros

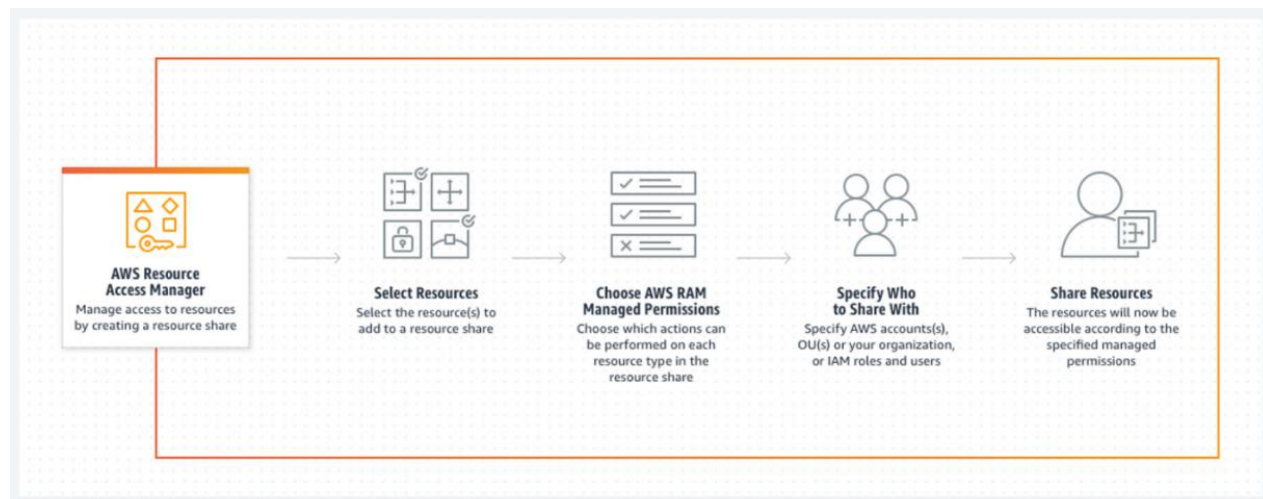
Com AWS RAM, você não precisa criar recursos duplicados em várias contas AWS. Isso reduz a sobrecarga operacional de gerenciamento de recursos em todas as contas que você possui

Em vez disso, em seu ambiente de várias contas, você pode criar um recurso uma vez e usar AWS RAM para compartilhar esse recurso entre contas, criando um compartilhamento de recurso

Ao criar um compartilhamento de recursos, você seleciona os recursos a serem compartilhados, escolhe uma permissão gerenciada do AWS RAM por tipo de recurso e especifica quem deseja que tenha acesso aos recursos.

Definição de Preço: A AWS RAM está disponível para você sem custo adicional

Como Funciona



AWS Systems Manager

O AWS Systems Manager é o hub de operações para a AWS. O Systems Manager oferece uma interface do usuário unificada para que você possa monitorar e solucionar problemas operacionais entre suas aplicações e recursos da AWS em um local central. **Com o Systems Manager, você pode automatizar tarefas operacionais para instâncias do Amazon EC2 ou do Amazon RDS**

O Systems Manager simplifica o gerenciamento de recursos e aplicações, reduz o tempo de detecção e solução de problemas operacionais e torna mais fácil a operação e o gerenciamento de infraestruturas em **grande escala**

Você pode usar runbooks de automação predefinidos ou criar os seus próprios para compartilhamento em tarefas operacionais comuns, como interromper e reiniciar uma instância do EC2

O Systems Manager apresenta os dados operacionais do grupo de recursos em um único e de fácil leitura, sem necessidade de navegar para outros Consoles AWS. Por exemplo, se você tiver um aplicativo que usa o Amazon EC2, o Amazon EKS, o Amazon S3 e o Amazon RDS, será possível usar o Systems Manager para criar um grupo de recursos para a aplicação e facilmente visualizar o software instalado nas suas instâncias do Amazon EC2, qualquer alteração em objetos do Amazon S3 ou as instâncias de banco de dados que foram interrompidas.

Trocando em Miúdos

É muito utilizado para executar scripts/comandos dentro de uma instância EC2. Requer um agente instalado no servidor. Ou seja, é muito utilizado para realizar automações, tanto automações já disponíveis no System Manager quanto automações customizadas

É possível tomar ações a nível de AWS quanto a nível da EC2

Ações a nível de AWS como por exemplo realizar um backup da instância, dar um stop/start na instância em um determinado horário dentre outras. Essas automações, que são conhecidas como Documentos, já estão pré configurados

Ações a nível da instância EC2, significa que eu posso chamar o meu agente que está dentro da instância e executar scripts (se for Linux, ShellScript, se for Windows, PowerShell), e o output da execução dos scripts podem ser vistos pelo console da AWS

É possível também restringir o acesso as instâncias via ssh por exemplo e permitir o acesso somente por uma console disponibilizada pelo System Manager. Pela console o usuário já é root e tem acesso irrestrito a execução de comandos dentro da instância. Todas as ações realizadas podem ser registradas e lançadas em um S3 e CloudWatch

Automação

O AWS Systems Manager permite automatizar com segurança tarefas comuns e repetitivas de operações e gerenciamento de TI

Com o Systems Manager Automation, você pode usar manuais predefinidos ou criar, executar e compartilhar manuais automatizados no estilo wiki para permitir o gerenciamento de recursos da AWS em várias contas e regiões da AWS

Agora, você pode executar scripts Python ou PowerShell como parte de um manual em combinação com outras ações de automação como aprovações, chamadas de APIs da AWS ou execução de comandos nas instâncias do EC2

Esses documentos podem ser agendados em uma janela de manutenção, acionados com base nas alterações aos recursos da AWS por meio do Amazon CloudWatch Events ou executados diretamente usando o Console de Gerenciamento da AWS, as CLIs e os SDKs

Você pode rastrear a execução de cada etapa em um manual, exigir aprovações, disponibilizar modificações de forma incremental e interromper automaticamente a disponibilização, caso ocorram erros

Definição de Preço: Gratuito

AWS Trusted Advisor (conselheiro de confiança)

O AWS Trusted Advisor faz **recomendações/sugestões** que ajudam você a seguir as melhores práticas da AWS. O Trusted Advisor avalia a sua conta através de verificações

Essas verificações identificam formas de otimizar sua infraestrutura da AWS, aumentar a segurança e o desempenho, reduzir os custos gerais e monitorar as cotas do serviço. Depois, você pode seguir as recomendações da verificação para otimizar seus recursos e serviços

Sugestões como:

- **Custos:** traz sugestões de como é possível melhorar os custos da conta
- **Performance:** por exemplo: recursos em excesso para determinado serviço
- **Segurança:** checa possíveis falhas de segurança e o que pode ser feito para ser melhorado, por exemplo: security group que é anexado a um recurso, dentro do security group pode haver regras de Allow ou Deny para portas e IPs, se o Trusted Advisor perceber que a porta 22 ssh está aberta para o mundo todo, será apresentado no painel como um problema de segurança
- **Tolerância a Falhas:** O Trusted Advisor pode aumentar a disponibilidade e a redundância de aplicativos da AWS, recomendando a utilização de recursos de autoscaling, verificações de integridade, regiões multi-AZ e backup.
- **Service Limits (*Soft limit):** As cotas de serviço, também chamadas de limites de serviço, são o número máximo de recursos de serviço ou operações aplicáveis a uma conta ou região. O Trusted Advisor pode notificar você se o uso da cota de serviço ultrapassar os 80%. Você pode, então, seguir as recomendações e excluir recursos ou solicitar um aumento de cota



Depende do Plano de Suporte que o usuário possui:

O Trust Advisor tem um engajamento direcionado para que o cliente possua um plano de suporte, dependendo do plano que ele assina, ele terá acesso completo as recomendações e pode acionar o suporte se necessário

AWS Basic Support e **AWS Developer Support:** podem acessar as principais verificações de segurança e todas verificações de Cotas de Serviço (Service Limits)

Clientes do **AWS Business Support** e do **AWS Enterprise Support:** podem acessar todas as verificações, incluindo otimizações de custo, segurança, tolerâncias a falhas, desempenho e cotas de serviço

Valores Baseados em Snapshots

Os valores são baseados em um snapshot, portanto a sua utilização atual pode ser diferente. Os dados de limite e de uso podem levar até 24 horas para refletir as alterações. Nos casos em que os limites foram aumentados recentemente, você continua a ver utilizações que excederam o limite por algum tempo.

* Soft Limits vs Hard Limits

Soft: mais flexível e pode-se aumentar o limite dos recursos (ativando o suporte da AWS), como por exemplo, escalonar mais de 100 instâncias EC2

Hard: mais engessado (não é possível aumentar o limite dos recursos ou das atividades mesmo ativando o suporte da AWS). O limite é EXCEED normalmente quando se tenta realizar algum tipo de procedimento (como por exemplo, fazer o snapshot de x servidores dentro de 1 segundo) utilizando a CLI ou SDK, inclusive o Console da AWS pode apresentar o erro, pois tem um SDK por trás. O que é comum como solução é os clientes criarem contas adicionais para provisionar mais recursos, como por exemplo, uma conta de DEV e uma de PROD

TAREFA: QUAL É O SOFT LIMIT DE RECURSOS COMO EC2, RDS, EBS, S3...

➤ **Pontos de Atenção**

Seu diretor de TI o instruiu a garantir que todos os recursos da AWS em seu VPC não ultrapassem o limite de serviço. Qual dos seguintes serviços pode ajudar nessa tarefa? R: AWS Trusted Advisor

AWS OpsWorks

Automatize as operações com o Chef e o Puppet

O AWS OpsWorks é um serviço de gerenciamento de configurações oferecem instâncias gerenciadas do Chef e do Puppet. O Chef e o Puppet são plataformas de automação que permitem usar código para automatizar a configuração de servidores. O OpsWorks permite usar o Chef e o Puppet para automatizar a forma como os servidores são configurados, implantados e gerenciados em instâncias do Amazon EC2 ou ambientes de computação no local. O OpsWorks tem três ofertas, AWS OpsWorks for Chef Automate, AWS OpsWorks for Puppet Enterprise e AWS OpsWorks Stacks.

What is Chef?

To quote the designers, Chef is “...a powerful automation platform that transforms infrastructure into code. Whether you’re operating in the cloud, on-premises, or in a hybrid environment, Chef automates how infrastructure is configured, deployed, and managed across your network, no matter its size.”

A chef is an open-source cloud configuration that translates system administration tasks into reusable definitions, otherwise known as cookbooks and recipes, which is why it’s named Chef. Makes sense, no?

Chef runs on a solid mix of platforms, including Windows; enterprise Linux distributions; AIX; FreeBSD; Solaris; Cisco IO; and Nexus. In addition, Chef also supports cloud platforms including Amazon Web Services (AWS), Google Cloud Platform, OpenStack, IBM Bluemix, HPE Cloud, Microsoft Azure, VMware vRealize Automation, and Rackspace.

What is Puppet?

According to the designer’s website, Puppet is an open-source systems management tool created for centralizing and automating configuration management. It includes its own declarative language to describe system configuration.

Puppet runs on the following platforms: Red Hat Enterprise Linux (and derivatives), SUSE Linux Enterprise Server, Debian, Ubuntu, Fedora, Microsoft Windows (Server OS), Microsoft Windows (Consumer OS) 10 Enterprise 7, 8, 10, macOS 10.12 Sierra, 10.13 High Sierra.

Puppet and its prerequisites have been reportedly run on the following platforms, though Puppet does not provide official open-source packages or perform automated testing:

Other Linux: Gentoo Linux, Mandriva Corporate Server 4, Arch Linux

Other Unix: Oracle Solaris version 10 and higher, AIX version 6.1 and higher, FreeBSD 4.7 and later, OpenBSD 4.1 and later, HP-UX.

➤ Pontos de Atenção

1. You are setting up a configuration management in your existing cloud architecture where you have to deploy and manage your EC2 instances including the other AWS resources using Chef and Puppet. Which of the following is the most suitable service to use in this scenario? R: AWS OpsWorks
2. Sua empresa requer um modelo baseado em pilha para seus recursos na AWS. É necessário ter pilhas diferentes para os ambientes de Desenvolvimento e Produção. Qual das seguintes opções pode ser usada para isso? R: Use AWS OpsWorks to define the different layers for your application. AWS OpsWorks Stacks lets you manage applications and servers on AWS and on-premises. *With OpsWorks Stacks, you can model your application as a stack containing different layers, such as load balancing, database, and application server. You can deploy and configure Amazon EC2 instances in each layer or connect other resources such as Amazon RDS databases.*